

THE UNIVERSITY OF MICHIGAN

DEPARTMENT OF ASTRONOMY
David M. Dennison Building

Ann Arbor, Michigan 48109-1090
Tel: (734)-764-3454/3440
jbregman@umich.edu

October 29, 2004

Alan Smale
NASA Headquarters
Office of Space Science
Code SZ
Washington, D.C. 20546-0001

Dear Alan,

The SAWG held a meeting on October 14-15, 2004 at NASA HQ, with the following members present: Julian Borrill, Joel Bregman (Chair), Roger Brissenden, Menas Kafatos, Carol Lonsdale, Tom McGlynn, Sally Oey, Rick White, and with new members Steve Murray, Gordon Squires, and Susan Neff, along with the NASA HQ personnel Alan Smale, Jeff Hayes, Paul Hertz, and Joe Bredekamp. Due to conflicts with the Senior Review and other demands upon Space Science personnel, there was not a meeting during the Spring, so this was the first meeting in 12 months. A number of small issues arose (presented first), but there were two major areas of discussion, one involving the Keck Observatory Archive and the other involving the Sloan Digital Sky Survey, both items dealing with NASA's relationship to ground-based observing. This common issue led to a separate letter that was sent to NASA in the event that it would be of interest to the Astronomy and Astrophysics Advisory Committee (AAAC), which was to meet during the week following the SAWG meeting.

GALEX

The SAWG heard an update on GALEX from David Schiminovich. GALEX was launched in April 2003, and officially started Guest Investigator observing in October 2004 (1/3 time). We heard that the first data release (GALEX Release 1, GR1) will occur at the end of October 2004, and the first GI data deliveries in January 2005, but the date of the second data release was not settled. The committee strongly recommends that GALEX should make annual public data releases. The committee was disappointed to hear that photon lists would not be part of GR1, and recommends that photon lists be included in all future data releases (in FITS format, as requested in a previous SAWG letter).

AISRP and other Computer Issues

Among the projects funded under the AISRP initiative, a fraction become widely used by the community (e.g., ds9), at which time there is a general need for such software to be maintained. However, the AISRP program has been a development initiative, not a maintenance initiative, so the process for maintenance of such software (and the role played by AISRP) is unclear, an issue discussed in our previous report. One example of the failings of the present situation is shown by the HEALPix software package, which has become a standard in the CMB

community, and an AISRP proposal for further improvement received an excellent rating but was rejected as inappropriate to the program. One possible solution is that software maintenance be assigned to the most relevant NASA data center for which the software would be a useful tool in dealing with the archival data. Such a transition would require additional funding to the center for maintenance and possible further development. Currently, the adoption of such software by centers has been on an ad-hoc basis, but the SAWG believes that this is not a sound policy and it should be replaced by a clear, well-articulated policy. The SAWG has raised this issue before, with little response, so we urge the relevant parties to develop such a policy soon and report on it at our next meeting.

Joe Bredekamp also reported on the recent deployment of the Columbia supercomputer at the Ames Research Center. We would like to commend all involved for the extraordinarily rapid deployment of this system, and congratulate the team for achieving the second highest performance of any supercomputer in the world in time. NASA now has an exceptional resource both for individual research programs and mission-critical data analysis.

LAMBDA and WMAP

Gary Hinshaw gave a presentation on the activities of the Legacy Archive for Microwave Background Data Analysis (LAMBDA), with much of the time being spent discussing the WMAP results and the processing of the 2-year WMAP data for general release. The WMAP data in the archive is the primary data set accessed by users, but as there has not been a new data release in 20 months, the SAWG tried to understand the activities of LAMBDA personnel, which could have been described more clearly. Since some LAMBDA staff also work for WMAP, there was concern as to whether the proscribed division between the two activities was being maintained. The support level for LAMBDA was predicated on the services and activities that they provide to the community, and those resources must be protected from the demands of WMAP mission analysis to avoid shortchanging the needs of the community.

There was a suggestion that there be a closer association between LAMBDA and the maintenance of the HEALPix software package, possibly in the form of financial support if appropriate. Also, another basic data product that could be added to LAMBDA is the first year unfiltered WMAP time-ordered data.

We are pleased to see that the LAMBDA users group had their inaugural meeting (a teleconference). There was concern that the composition of the users group did not have adequate representation of core US users of CMB data and we hope that this shortcoming will be corrected before the next users group meeting.

INTEGRAL

The SAWG recommends that INTEGRAL pursue a joint plan with ESA to develop a concrete plan and timeframe for the inclusion of Level-1 products in the data archive. We further recommend proceeding towards providing data that is compatible with the standard FTOOLS package, insofar as that is possible.

The Michelson Science Center and Keck Observatory Archive

The SAWG appreciated David Ciardi's introduction to the Michelson Science Center (MSC), which is sited at Caltech, within IPAC where the MSC takes advantage of IRSA's data architectures and scientific experience. The MSC is chartered to provide data analysis tools for the science teams of the interferometry missions, and to provide an interface from the missions to

the broader scientific community. Missions currently included in the MSC charter are the Keck interferometer, the Large Binocular Telescope Interferometer (not all LBT data), the Palomar Testbed Interferometer, and the Keck HIRES spectra. The MSC is currently archiving and serving the extensive ground-based pre-cursor data needed to plan the science programs for future missions including SIM and TPF. The MSC has begun serving Keck Interferometer and HIRES data to individual PI's, and will begin serving this data to the public in 2005 April (Interferometer) and 2006 February (HIRES). The 18 month proprietary interval seems longer than most other archives, though the committee applauds the efforts to enable to the public release of these data.

A subsequent briefing by Carl Pilcher brought up concerns that the Keck Observatory Archive (KAO) has in developing a policy for proprietary data for projects that may require many years of data, notably long term monitoring of stars for planets. Carl discussed several options that have been considered for addressing this, and requested the SAWGs feedback on the proprietary policy. Carl provided an augmentation plan for the Keck archive which included options for a Level 1 pipeline.

The committee had several concerns with these presentations:

- 1) We strongly encourage the MSC to adopt existing data formats for astronomical interferometry data. Where possible, the MSC should use existing standards developed in earlier work in interferometry at other wavelengths, extending them as needed to accommodate the differences between phase and amplitude measurements and other special features of optical interferometry. Providing data in formats that work in established community tools will greatly enhance the ability of non-specialists to use the data in this field. The MSC should work with other groups in optical interferometry to establish these standards throughout the field.
- 2) The MSC is currently directed by NASA HQ to archive *only* Level 0 data; however we feel it desirable that data served by the archive should be available at both Level 0 *and* Level 1, and experience at other archives shows that most users retrieve Level 1 data. The KOA expressed a desire to build a pipeline for Level 1 data, and that this would be a relatively inexpensive undertaking. There was strong concern by the SAWG that the MSC has severely underestimated the effort to build a Level 1 pipeline. The regular production of Level 1 data sets will require a different method of observing, and the sweeping implications for this cultural change needs to be considered by the MSC. The development of a pipeline for Level 1 data is best accomplished as a shared activity with the other (primary) members of the Keck Observatory.
- 3) We encountered questions about NASA archiving ground-based data with respect to the Keck HIRES data. About 30% of HIRES data is directly connected to the goal of finding extrasolar planets, and these are the data that are archived, by agreement. The MSC would like to archive the remaining 70%, and while generally laudable, this raises the concern that NASA is funding archives for ground data which should be supported by other institutions. We recommend that extensions of the MSC archiving to other Keck data types be carefully considered in light of a more general strategy for NASA support of ground data. The SAWG believes that the other partners in the observatory share ownership of the Keck archive and we recommend that a joint approach is most desirable.

4) A flexible proprietary period for long-term projects using HIRES data seems appropriate. Generally the SAWG favors policies where there is some release of data on timescales comparable to the normal proprietary period (most commonly, 1 year). The policy should attempt to maximize the usability of the data for science programs unrelated to the long-term goals of the PI. The suggested rollout scenarios seemed reasonable.

Sloan Digital Sky Survey

Richard Kron (University of Chicago) and Alex Salay (JHU) presented the status and plans for the Sloan Digital Sky Survey (SDSS) archive project. The Project is planning on a 3 year, \$15M, extension to the Survey with funding expected from NSF, and matching funding from the Sloan Foundation and other Institutional commitments. The science goals for the extension include a Legacy survey of the north Galactic cap, the SEGUE (Sloan Extension for Galactic Underpinnings and Evolution) survey to study Galactic structure and stellar astrophysics including stellar abundance measurements, and a search for supernovae in the undersampled regime $0.1 < z < 0.3$.

The SDSS archive and associated tools are well-conceived and impressive. The SDSS team has incorporated a cutting edge data base design approach involving automated schema generation and self-documentation, and have worked closely with the NSF-funded Virtual Observatory (VO) team to ensure that the data, metadata and services are suitable for use as part of the VO. The data base is therefore valuable as a tool for multiwavelength correlative astronomical research. At present the archive is being maintained by the Fermi Laboratory, but the commitment is not open-ended and there is a serious concern as to how the archive would be maintained following the planned 3-year extension in 2008.

The SAWG sees the SDSS archive as an important resource for astronomical research both as a standalone data set and as part of the VO. We recognize that the SDSS archive is a testbed for the interagency VO efforts. The data and data base are of outstanding quality and value to the science community, so a plan should be developed to ensure the long-term curation of the SDSS archive. Since the data are ground-based and have been produced with significant support from NSF and the DOE, it would be appropriate for one or both of these organizations to provide the long-term support for the data set, with advisory input from NASA. A cooperative effort between agencies is not only desirable, but necessary for future endeavors (e.g., a VO) and cooperative efforts can be organized through cross-agency committees such as the AAAC.

The issues surrounding the KOA and the SDSS, while very different, require a policy on the archiving of ground-based data. While NASA has a general policy on these matters, it may not be sufficiently detailed to deal with some of the issues that now arise. The SAWG would like more detailed guidance on these matters, as well as an understanding with other relevant agencies (NSF, DOE) as to their role in archiving data. This is especially important not only in light of future surveys, but in recognition that the vast majority of astronomical data is collected by ground-based observatories.

Sincerely yours,

Joel N. Bregman
Chair, SAWG